Verständnisorientierter gymnasialer Stochastikunterricht – quo vadis?

NORBERT HENZE, KARLSRUHE

Zusammenfassung: *Der gymnasiale Stochastikunterricht ist momentan von einer starken Rezeptorientierung, einer Priorisierung von (oft vermeintlichem) Anwendungsbezug und einem Mangel an Begründungen geprägt. Wir nehmen eine Bestandsaufnahme anhand wichtiger Grundbegriffe vor und zeigen Möglichkeiten für ein Entgegenwirken insbesondere angesichts einer zukünftigen Differenzierung in Grund- und Leistungskurse auf.*

1 Einleitung

Ein Blick in neuere Mathematikbücher wie z. B. Freudigmann et al. (2016) offenbart das Dilemma des Stochastikunterrichts an Gymnasien, vermutlich nicht nur in Baden–Württemberg. Festzustellen ist

- eine starke Rezeptorientierung,
- ein übertriebener (vermeintlicher) Anwendungsbezug,
- ein schematisches Testen des Parameters p der Binomialverteilung anhand meist „erfundener" Hypothesen,
- ein eklatanter Mangel an Begründungen,
- ein prinzipielles Defizit, was die Förderung begrifflichen Verständnisses betrifft.

Die Gründe hierfür sind sicherlich vielschichtig. Sie hängen nicht nur mit der Schulzeitverkürzung auf acht Jahre zusammen, sondern wohl auch damit, dass viele Gymnasiallehrerinnen und -lehrer im Studium entweder gar keine Stochastikvorlesung gehört haben oder ihnen eine Einführung in die Stochastik vermittelt wurde, die Priorität auf Maß- und Wahrscheinlichkeitstheorie setzte und – wenn überhaupt – wenig Bezug zur schulischen Stochastik hatte. Zudem begünstigt auch der aktuelle Bildungsplan des Gymnasiums Baden–Württemberg im Fach Mathematik vom 23.3.2016 eine Orientierung an Kochrezepten, wenn es etwa unter der Leitidee *Daten und Zufall* für die Klassenstufen 9 und 10 darum geht, „bei Binomialverteilungen den jeweils fehlenden Parameter (n, p oder k) mit geeigneten Hilfsmitteln zu bestimmen" oder „die Kenngrößen Erwartungswert und Standardabweichung einer binomialverteilten Zufallsgröße zu berechnen".

In diesem Aufsatz, dem ein Vortrag anlässlich der Jahrestagung 2017 des Arbeitskreises Stochastik der Gesellschaft für Didaktik der Mathematik zugrunde liegt, nehmen wir eine Bestandsaufnahme von prinzipiellen Defiziten anhand der Grundbegriffe *Zufallsvariable, Erwartungswert, Binomialkoeffizient, binomische Formel, Bernoulli-Kette, Binomialverteilung, Varianz sowie Standardabweichung* vor. Zudem zeigen wir auf, wie ein auf größeres begriffliches Verständnis und eine stärkere Gewichtung von Begründungen zielender Unterricht insbesondere angesichts einer geplanten Differenzierung in Grund- und Leistungskurse aussehen könnte.

2 Zufallsvariablen – das, was uns jeweils interessiert?

In Bostelmann et al. (2009) kann man auf Seite 63 lesen:

> „Spielst Du z.B. Monopoly, dann kommt es auf die Augensumme an, bei anderen Spielen auf den Unterschied der beiden Augenzahlen. Wir nennen die Größe, auf die es uns ankommt, Zufallsvariable."

Beim „CHECK UP" auf Seite 92 heißt es dann noch einmal zusammenfassend:

> „**Zufallsvariable** X**:** Die Größe, auf die es uns ankommt (hier der Gewinn)."

Welch ungeheure konzeptionelle und intellektuelle Abrüstung mit solchen Formulierungen verbunden ist, kann man daran ersehen, dass die Vorstellung von einer Zufallsvariablen X als einer auf der Ergebnismenge Ω eines stochastischen Vorgangs definierten reellen Funktion noch klar in Barth und Haller (1985, S.165) herausgearbeitet und die Zuordnung von $\omega \in \Omega$ zu $X(\omega)$ durch Pfeile veranschaulicht wurde. Später ging man wie etwa in Felmy et al. (1986, S. 74), dazu über, Zufallsvariablen an einem Beispiel einzuführen, aber zumindest noch (beiläufig) zu sagen, dass die Zufallsvariable *jedem Ergebnis eine Zahl zuordnet*. Problematischer wird es jedoch schon auf Seite 108 in Biehler et al. (2012), wo eine Zufallsgröße als

> „Variable [verstanden wird, die] als Wert eine der betreffenden reellen Zahlen annimmt, je nachdem, wie der Versuch ausgeht."

Hier wird der Blick zu schnell vom Grundraum weggelenkt, was insbesondere dem Umgang mit Erwartungswerten abträglich ist.

Steht $\Omega = \{\omega_1, \ldots, \omega_k\}$ für die Menge der Ergebnisse eines stochastischen Vorgangs, so ist eine (reelle) *Zufallsvariable* eine Abbildung, die jedem $\omega \in \Omega$ eine reelle Zahl $X(\omega)$, die sogenannte *Realisierung von X zum Ausgang* Ω, zuordnet. Diese Vorstellung von einer Zufallsvariablen als Zuordnung, und zwar von Ergebnissen des stochastischen Vorgangs zu reellen Zahlen, ist grundlegend für das Verständnis des Begriffs Zufallsvariable.

Als Beispiel betrachten wir den Grundraum $\Omega = \{(i,j) : i,j \in \{1,2,3,4,5,6\}\}$ als Modell für den zweimal hintereinander ausgeführten Würfelwurf. Anhand der durch

- $X(i,j) := i + j$,
- $Y(i,j) := \max(i,j)$,
- $Z_1(i,j) := i, \quad Z_2(i,j) := j$

definierten Zufallsvariablen X, Y, Z_1 und Z_2 wird deutlich, dass bei einem stochastischen Vorgang unterschiedliche Aspekte wichtig sein können. Insbesondere sieht man, dass die Augensumme X die *Summe der Zufallsvariablen* Z_1 *und* Z_2 ist.

Hat man nämlich das Konzept einer Zufallsvariablen als Zuordnung verinnerlicht, so ist unmittelbar klar, dass man mit Zufallsvariablen X und Y und einer reellen Zahl a „rechnen“ kann, indem man etwa für jedes $\omega \in \Omega$

$$\begin{aligned}
(X \pm Y)(\omega) &:= X(\omega) \pm Y(\omega), \\
(X \cdot Y)(\omega) &:= X(\omega) \cdot Y(\omega), \\
(a \cdot X)(\omega) &:= a \cdot X(\omega), \\
(\max(X,Y))(\omega) &:= \max(X(\omega), Y(\omega))
\end{aligned}$$

setzt. Man sollte auch nie einen wichtigen Zweck von Zufallsvariablen vergessen: Sie dienen der bequemen Beschreibung von Ereignissen wie etwa

$$\begin{aligned}
\{X = \ell\} &:= \{\omega \in \Omega : X(\omega) = \ell\}, \\
\{X \leq \ell\} &:= \{\omega \in \Omega : X(\omega) \leq \ell\} \quad \text{usw.}
\end{aligned}$$

Die wichtigste nicht konstante Zufallsvariable nimmt nur zwei Werte an: Ist $A \subseteq \Omega$ ein Ereignis, so heißt die durch

$$\mathbf{1}_A(\omega) := \begin{cases} 1, & \text{falls } \omega \in A, \\ 0, & \text{falls } \omega \notin A, \end{cases}$$

definierte Zufallsvariable $\mathbf{1}_A : \Omega \to \mathbb{R}$ *Indikatorfunktion von A* oder kurz *Indikator* von A. Die Realisierung dieser Zufallsvariablen zeigt an, ob das Ereignis A eingetreten ist oder nicht.

Sind $A_1, \ldots, A_n \subseteq \Omega$ Ereignisse, so modelliert die sogenannte *Indikatorsumme*

$$X := \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \ldots + \mathbf{1}_{A_n} \tag{1}$$

die Anzahl der eintretenden Ereignisse unter $A_1, \ldots, A_n$. In der Tat nimmt X die möglichen Werte $0, 1, \ldots, n$ an, und es gilt $X(\omega) = k$, wenn ω zu genau k der Ereignisse $A_1, \ldots, A_n$ gehört. Beispiele für Indikatorsummen sind die Anzahl der Treffer in einer Bernoulli-Kette, die Anzahl der gezogenen roten Kugeln beim Ziehen ohne Zurücklegen, die Anzahl der Fixpunkte einer zufälligen Vertauschung von Zahlen, die Anzahl der freien Fächer beim Verteilen von Teilchen auf Fächer oder die Anzahl der Mehrfachgeburtstage unter n Personen.

3 Der Erwartungswert – nur noch „Ausrechnen nach Rezept“?

Der Erwartungswert einer Zufallsvariablen X wird wie etwa auf Seite 133 in Baum et al. (2007) am Beispiel eines „bei einer großen Anzahl von Durchführungen des Zufallsversuchs durchschnittlich zu erwartenden Wertes für X“ eingeführt. Danach kommt sofort das „Summe Wert mal Wahrscheinlichkeit–Rezept“

$$\mathbb{E}(X) = \sum_{i=1}^{s} x_i \cdot \mathbb{P}(X = x_i) \tag{2}$$

zur Berechnung von $\mathbb{E}(X)$, wenn die Zufallsvariable X die Werte $x_1, \ldots, x_s$ annimmt. Besitzt X speziell eine Binomialverteilung mit Parametern n und p, so erfährt man, dass $\mathbb{E}(X) = n \cdot p$ gilt, siehe z. B. Freudigmann et al. (2016, S. 140). Da der Nachweis dieser Gleichung über obiges Rezept auf das Problem führt, die Summe

$$\mathbb{E}(X) = \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \tag{3}$$

zu berechnen und der allgemeine binomische Lehrsatz nicht (mehr) bekannt ist, wird dieser Nachweis nur im Spezialfall $n = 3$ geführt. Hoffentlich gibt es noch Schülerinnen und Schüler, die sich nicht damit zufrieden geben, siehe hierzu Abschnitt 5.

In meiner Vorlesung *Einführung in die Stochastik für Studierende des gymnasialen Lehramts Mathematik* (Henze, o. D.) führe ich den Erwartungswert einer Zufallsvariablen X wie folgt ein: Ist $\Omega = \{\omega_1, \ldots, \omega_k\}$ die Menge der Ergebnisse eines stochastischen Vorgangs, so gibt der Wert $X(\omega_j)$ die Realisierung der Zufallsvariablen X beim Ausgang ω_j dieses Vorgangs an. Führt man den Vorgang n-mal durch, und tritt hierbei der Ausgang ω_j h_j mal auf, $j = 1, \ldots, k$, so ergibt sich als Gesamtsumme der Realisierungen von X der Wert

$$X(\omega_1) \cdot h_1 + \ldots + X(\omega_k) \cdot h_k.$$

Der *durchschnittliche Wert pro Vorgang* – in einer üblichen Glücksspiel-Einkleidung also der durchschnittliche Gewinn pro Spiel – ist dann

$$X(\omega_1) \cdot \frac{h_1}{n} + \ldots + X(\omega_k) \cdot \frac{h_k}{n}. \tag{4}$$

Da sich die relativen Häufigkeiten h_j/n der einzelnen Ausgänge ω_j unter gleichen, „unabhängigen Bedingungen" bei wachsendem n gegen die Wahrscheinlichkeiten $\mathbb{P}(\{\omega_j\})$ stabilisieren sollten, ist die folgende Definition selbstredend:

Definition 1: Die Zahl

$$\mathbb{E}(X) := \sum_{j=1}^{k} X(\omega_j) \cdot \mathbb{P}(\{\omega_j\}) \tag{5}$$

heißt *Erwartungswert* von X.

Aus dieser noch auf Seite 172 in Barth und Haller (1985) sogar in der abstrakteren Form

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

aufgeführten Darstellung ergibt sich durch Zusammenfassen von möglicherweise gleichen Werten $X(\omega_j)$ die wohlbekannte Gleichung (2). Man erhält aber auch die folgenden, unschätzbaren (weil manchmal das Rechnen durch Denken ersetzenden) Einsichten für den Umgang mit Erwartungswerten:

Satz 1: Für Zufallsvariablen X und Y gelten:

a) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,

b) $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}(X), \quad a \in \mathbb{R}$,

c) $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A), \quad A \subseteq \Omega$,

d) Aus $X \leq Y$ folgt $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Dabei besagt die Ungleichung $X \leq Y$, dass $X(\omega_j) \leq Y(\omega_j)$ für jedes $j = 1, \ldots, k$ zutrifft. Aus a) und c) ergibt sich, dass der Erwartungswert einer Indikatorsumme $\mathbf{1}_{A_1} + \ldots + \mathbf{1}_{A_n}$ durch

$$\mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{A_j}\right) = \sum_{j=1}^{n} \mathbb{P}(A_j) \tag{6}$$

gegeben ist. Ist die Zufallsvariable X also von der Gestalt (1), so benötigt man nicht die *Verteilung* von X, um $\mathbb{E}(X)$ zu erhalten, sondern nur die Wahrscheinlichkeiten von $A_1, \ldots, A_n$. Im wichtigen Spezialfall, dass $A_1, \ldots, A_n$ die gleiche, mit p bezeichnete Wahrscheinlichkeit besitzen (aber keinen weiteren Einschränkungen wie z. B. stochastische Unabhängigkeit genügen müssen), folgt

$$\mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{A_j}\right) = n \cdot p. \tag{7}$$

Betrachtet man (7), so fällt es wie Schuppen von den Augen, *warum* der Erwartungswert einer binomialverteilten Zufallsvariablen mit Parametern n und p gleich $n \cdot p$ ist, denn X zählt ja, wie viele von n Ereignissen $A_1, \ldots, A_n$, die alle die gleiche Wahrscheinlichkeit p besitzen, eintreten. Dabei deutet man in diesem Zusammenhang das Eintreten des Ereignisses A_j als Treffer im j-ten Versuch, $j = 1, \ldots, n$.

Wir möchten jetzt anhand zweier Beispiele zeigen, welche Einsichten für den Unterricht in einer Kursstufe des Gymnasiums allein durch Einführung des Erwartungswertes über die Darstellung (5) gewonnen werden können.

Beispiel 1: Es sei

$$\Omega = \left\{ \begin{matrix} 1234 & 2134 & 3124 & 4123 \\ 1243 & 2143 & 3142 & 4132 \\ 1324 & 2314 & 3214 & 4213 \\ 1342 & 2341 & 3241 & 4231 \\ 1423 & 2413 & 3412 & 4312 \\ 1432 & 2431 & 3421 & 4321 \end{matrix} \right\}$$

die Menge aller 24 möglichen Vertauschungen (Permutationen) der Zahlen 1 bis 4, wobei jede Vertauschung gleich wahrscheinlich sei. Von links nach rechts gelesen sehen wir eine Zahl als *Rekord* an, wenn sie größer als alle (eventuell) davor stehenden Zahlen ist. Formal betrachten wir somit für jedes $j = 1,2,3,4$ das Ereignis

$$A_j := \{a_1 a_2 a_3 a_4 \in \Omega : a_j = \max(a_1, \ldots, a_j)\}. \tag{8}$$

Direktes Abzählen der jeweils günstigen unter allen 24 Vertauschungen liefert

$$\mathbb{P}(A_1) = 1,\ \mathbb{P}(A_2) = \frac{1}{2},\ \mathbb{P}(A_3) = \frac{1}{3},\ \mathbb{P}(A_4) = \frac{1}{4},$$

sodass Gleichung (6) das Resultat

$$\mathbb{E}\left(\sum_{j=1}^{4} \mathbf{1}_{A_j}\right) = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{25}{12} = 2{,}0833\cdots$$

liefert. Schreiben wir kurz X für die Anzahl der Rekorde, so lässt sich natürlich auch durch einfaches Abzählen die Verteilung von X zu

j	1	2	3	4
$\mathbb{P}(X=j)$	$\frac{6}{24}$	$\frac{11}{24}$	$\frac{6}{24}$	$\frac{1}{24}$

ermitteln, und man bestätigt durch Anwenden des „Rezeptes" (2) das obige Ergebnis. Pfiffige Schülerinnen und Schüler könnten die Frage stellen, was allgemein bei der Vertauschung von n Zahlen passiert und vielleicht sogar selbst die Antwort geben.

Vertauscht man n Zahlen rein zufällig, so ist – ganz gleich, welche dieser Zahlen von links nach rechts gelesen die ersten j Zahlen bilden – genau eine davon die größte. Die Wahrscheinlichkeit, dass diese größte Zahl an der j-ten Stelle steht, muss aus Symmetriegründen $1/j$ sein. Die Anzahl der Rekorde bei einer rein zufälligen Vertauschung von n Zahlen besitzt also den Erwartungswert

$$\mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{A_j}\right) = \sum_{j=1}^{n} \frac{1}{j}.$$

Dabei bezeichnet A_j wie oben (nur nach Ersetzen von 4 durch n und Modifikation von Ω) das Ereignis, dass an der j-ten Stelle ein Rekord auftritt. Wegen

$$\begin{aligned}\ln(n+1) &= \int_0^n \frac{\mathrm{d}x}{1+x} \leq \sum_{j=1}^{n} \frac{1}{j} \\ &\leq 1 + \int_1^n \frac{\mathrm{d}x}{x} = 1 + \ln n\end{aligned}$$

erkennt man, dass die Anzahl der Rekorde recht langsam, nämlich logarithmisch, mit n wächst. Es sollte auch klar sein, dass es *irgendwelche* verschiedenen Zahlen sein können, die rein zufällig vertauscht werden wie etwa jährliche Temperaturmittelwerte. Gäbe es keinen Klimawandel, so sollten alle Vertauschungen die gleiche Wahrscheinlichkeit besitzen. Gibt es zu viele Rekorde, so würde man die Annahme der Gleichverteilung über alle Vertauschungen verwerfen. Hierzu muss man natürlich Informationen über die Verteilung der Anzahl der Rekorde besitzen, siehe z. B. Henze (2008).

Beispiel 2: Wir legen wie in Beispiel 1 die Menge Ω aller Permutationen der Zahlen 1 bis 4 zugrunde, interessieren uns aber jetzt für eine andere Zufallsvariable, nämlich die Anzahl der Fixpunkte. Dabei ist formal das Ereignis „j ist Fixpunkt“ für jedes $j = 1,2,3,4$ durch

$$A_j := \{a_1a_2a_3a_4 \in \Omega : a_j = j\}$$

definiert. Direktes Betrachten der jeweils günstigen Fälle ergibt

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \mathbb{P}(A_4) = \frac{1}{4},$$

und mit (7) folgt

$$\mathbb{E}\left(\sum_{j=1}^{4} \mathbf{1}_{A_j}\right) = 4 \cdot \frac{1}{4} = 1.$$

Im Mittel gibt es also genau einen Fixpunkt. Auch hier erkennen pfiffige Schülerinnen und Schüler sofort, was allgemein bei einer rein zufälligen Vertauschung der Zahlen $1,2,\ldots,n$ passiert. Es gibt $(n-1)!$ Permutationen, die an der j-ten Stelle den Fixpunkt j besitzen, denn die anderen $n-1$ Zahlen sind ja beliebig vertauschbar. Also ist (wieder nach Ersetzen von 4 durch n) $\mathbb{P}(A_j) = (n-1)!/n! = 1/n$, und man erhält mit (7), dass unabhängig von n im Mittel genau ein Fixpunkt auftritt.

Für die *Verteilung* der Anzahl X der Fixpunkte gilt (s. z. B. Henze (2017, S. 75))

$$\mathbb{P}(X = k) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}, \qquad k = 0,1,\ldots,n.$$

4 Binomialkoeffizienten – hilflos ohne GTR?

Welches Zerrbild der Mathematikunterricht selbst bei den für das Verständnis von Anzahlen grundlegenden Binomialkoeffizienten abgeben kann, zeigt die Einführung der Binomialkoeffizienten auf Seite 84 in Bostelmann et al. (2009). Dort wird die Frage aufgeworfen, wie viele Pfade in einer Bernoulli-Kette der Länge 12 mit genau 8 Treffern existieren. Die dazu gegebene Antwort lautet wie folgt:

> „Zum Berechnen der betreffenden Anzahl gibt es eine mathematische Formel, den Binomialkoeffizienten, den man mithilfe des Taschenrechners auswerten kann. Wir schreiben $\binom{12}{8}$ und sagen *12 über 8*. $\binom{n}{r}$ steht also für die Pfade der Länge n mit genau r Treffern. Für das Beispiel erhalten wir mit dem GTR stolze 495 verschiedene Pfade.“

In diesem Buch wird also nicht einmal die definierende Gleichung

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} \tag{9}$$

hingeschrieben, sondern sofort zum Ausrechnen von „n über k“ auf den grafikfähigen Taschenrechner verwiesen. Die verheerende Botschaft lautet: Drücke auf die richtigen Tasten und frage nicht nach dem Warum!

Zum Glück wird Gleichung (9) in anderen Büchern wie z. B. Freudigmann et al. (2016, S. 134) noch begrifflich durch einen Abzählvorgang hergeleitet, sodass der Binomialkoeffizient $\binom{n}{k}$ als Anzahl der Möglichkeiten, aus n Objekten k Objekte ohne Berücksichtigung der Reihenfolge auszuwählen, verstanden wird. Im gleichen Buch wird auf Seite 154 die Frage aufgeworfen, warum die Zahlen im Pascalschen Dreieck mit den Binomialkoeffizienten übereinstimmen. Diese Frage läuft also darauf hinaus, die Gültigkeit der Rekursionsformel

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \quad 1 \leq k \leq n, \tag{10}$$

zu zeigen. Auf Seite 154 findet man den Beweis von (10) mithilfe direkter Rechnung mit den die Binomialkoeffizienten bildenden Fakultäten. Es wird jedoch keinerlei begriffliche Einsicht vermittelt, *warum* diese Formel gilt, dabei fällt sie doch, nachdem man die obige Bedeutung des Binomialkoeffizienten verstanden hat, als Nebenprodukt ab (und lässt sich damit ganz leicht merken):

Man nummeriert die Objekte von 1 bis n durch und unterscheidet alle k-Auswahlen der n Objekte danach, ob sie das Objekt n enthalten oder nicht. Im ersten Fall muss man noch $k-1$ Objekte aus den Objekten $1,2,\ldots,n-1$ auswählen, was auf $\binom{n-1}{k-1}$ Weisen möglich ist. Im zweiten Fall sind noch k Objekte aus den Objekten $1,2,\ldots,n-1$ auszuwählen, und hierfür gibt es $\binom{n-1}{k}$ Möglichkeiten.

Warum gönnt man wie früher (siehe z. B. Glaser et al. (1990, S. 45)) zumindest den interessierteren Schülerinnen und Schülern nicht die Freude, die Gültigkeit der Rekursionsformel auf diese Weise zu begreifen und zu erfahren, dass man Rechnen durch Denken ersetzen kann?

5 Die binomische Formel – fast eine Übungsaufgabe?

Die nicht nur im Zusammenhang mit der Binomialverteilung wichtige binomische Formel

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k} \qquad (11)$$

scheint – zumindest suggeriert dies Dürr et al. (2017) – selbst aus dem Unterricht der Kursstufe verschwunden zu sein. Dabei erschließt sich diese Gleichung unmittelbar als Nebenprodukt aus dem Verständnis der Binomialkoeffizienten als Anzahlen von Auswahlen: Um $(a+b)^n$ auszurechnen, schreibt man die Potenz als Produkt mit n Faktoren in folgender Form aus:

$$(a+b)\cdot(a+b)\cdot(a+b)\cdot\ldots\cdot(a+b)\;.$$

Nach den Regeln der Klammerrechnung ist aus jeder Klammer entweder a oder b auszuwählen und das Produkt der gewählten Zahlen zu bilden. Hat man sich k mal für a (und damit $n-k$ mal für b) entschieden, so entsteht als Produkt $a^k \cdot b^{n-k}$. Dieses tritt nach der Erkenntnis über den Binomialkoeffizienten genau $\binom{n}{k}$ mal auf, denn so oft kann man aus den n Klammern (Objekten) k hierfür auswählen. Da man zu guter Letzt alle $\binom{n}{k}\cdot a^k \cdot b^{n-k}$ über k von 0 bis n summieren muss, ist (11) bewiesen. Diese Formel ist mit etwas Denkaufwand so selbstverständlich, dass sie z. B. noch in Glaser et al. (1990) auf Seite 42 auftritt und in Barth und Haller (1985) auf Seite 115 (mit dem Hinweis: Überlege, wie oft der Summand $a^k b^{n-k}$ bei der Multiplikation entsteht?) sogar als *Übungsaufgabe* formuliert war!

Mit der binomischen Formel erhält man relativ leicht den Erwartungswert der Binomialverteilung über das „Erwartungswertbildung-Rezept“: Hierzu muss man zunächst einsehen, dass die Summation in (3) von $k=1$ ab erfolgen kann und dass für solche k

$$k\cdot\binom{n}{k} = n\cdot\binom{n-1}{k-1}$$

gilt. Spaltet man aus der Summe den Faktor p ab und schreibt $n-k$ als $n-1-(k-1)$, so folgt

$$\mathbb{E}(X) = n\cdot p\sum_{k=1}^{n}\binom{n-1}{k-1}\cdot p^{k-1}(1-p)^{n-1-(k-1)}.$$

Wenn man jetzt $k-1$ zu j abkürzt und bedenkt, dass sich die Summation über j von 0 bis $n-1$ erstreckt, so ergibt sich wegen

$$\sum_{j=0}^{n-1}\binom{n-1}{j}\cdot p^j(1-p)^{n-1-j} = (p+1-p)^{n-1} = 1$$

das Resultat $\mathbb{E}(X) = n\cdot p$ nicht nur im Spezialfall $n=3$, sondern allgemein.

6 Die Bernoulli-Kette – Vorsicht mit zu einfachen Definitionen!

Auf Seite 82 in Bostelmann et al. (2009) ist zu lesen:

> „Wird ein Zufallsversuch mehrfach wiederholt, dann nennt man diese Folge von Zufallsversuchen eine *Bernoulli-Kette*, wenn:
>
> - jeder Versuch zwei Ergebnisse hat: Treffer (T) oder Fehlschlag (F),
> - die Versuchswiederholungen unabhängig sind, d.h. die Wahrscheinlichkeiten p für Treffer und q für Fehlschlag bleiben auf jeder „Stufe“ gleich.“

Abgesehen davon, dass jeder Versuch nicht zwei, sondern nur ein Ergebnis hat (er hat zwei *mögliche* Ergebnisse), werden wir jetzt zeigen, dass das n-malige Ziehen *ohne* Zurücklegen aus einer Urne mit roten und schwarzen Kugeln in obigem Sinn eine Bernoulli-Kette darstellt.

Wir betrachten hierzu der Einfachheit halber eine Urne mit vier Kugeln, die – was den Zufall nicht beeinflusst – von 1 bis 4 nummeriert sind. Die Kugeln mit den Nummern 1 und 2 färben wir rot, die anderen Kugeln schwarz. Das Ziehen einer roten bzw. schwarzen Kugel bezeichnen wir als Treffer bzw. Fehlschlag (Niete). Ziehen wir gedanklich blind alle vier Kugeln nacheinander, so liegt die schon früher verwendete Ergebnismenge

$$\Omega = \begin{Bmatrix} 1234 & 2134 & 3124 & 4123 \\ 1243 & 2143 & 3142 & 4132 \\ 1324 & 2314 & 3214 & 4213 \\ 1342 & 2341 & 3241 & 4231 \\ 1423 & 2413 & 3412 & 4312 \\ 1432 & 2431 & 3421 & 4321 \end{Bmatrix}$$

vor. Das Ereignis A_j, im j-ten Zug eine rote Kugel zu ziehen, besteht dann aus denjenigen der 24 möglichen Ergebnisse $a_1a_2a_3a_4$, für die $a_j = 1$ oder $a_j = 2$ gilt. Jede Schülerin und jeder Schüler zählt sofort ab, dass es für das Eintreten von A_j unabhängig von j jeweils 12 günstige Fälle gibt. Aufgrund der Laplace-Annahme gilt also

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \mathbb{P}(A_4) = \frac{1}{2},$$

was zeigt, dass sich die Trefferwahrscheinlichkeit von Stufe zu Stufe *nicht* ändert.

Ein analoger Sachverhalt gilt natürlich auch allgemeiner beim n-fachen rein zufälligen Ziehen *ohne* Zurücklegen aus einer Urne mit r roten und s schwarzen Kugeln, denn auch beim blinden Ziehen *ohne* Zurücklegen hat jede der $r+s$ Kugeln die gleiche Chance, als j-te gezogen zu werden. Die Wahrscheinlichkeit, dass die j-te gezogene Kugel rot ist, ist also unabhängig von j gleich $r/(r+s)$.

Freudigmann et al. (2016) merken auf Seite 130 richtigerweise an, dass das Ziehen ohne Zurücklegen *keine* Bernoulli-Kette darstelle, aber mit der irreführenden Formulierung, dass sich bei jedem Zug die Trefferwahrscheinlichkeit ändern würde. Was aber tatsachlich gezeigt wird, ist, dass die *bedingten Wahrscheinlichkeiten* für einen Treffer im zweiten Zug davon abhängen, ob im ersten Zug ein Treffer auftritt oder nicht. Wenn man die zu Beginn dieses Abschnittes gemachte Formulierung über die „Unabhängigkeit der Versuchswiederholungen" dahin fortsetzen würde, dass auf jeder Stufe die Wahrscheinlichkeit p für Treffer *nicht von den Ergebnissen früherer Stufen abhängt*, wäre alles in Ordnung, weil dann Ereignisse, die sich auf verschiedene Stufen beziehen, *stochastisch unabhängig* sind.

In Freudigmann et al. (2016, S. 130) wird ein *Bernoulli-Experiment* als

> „*Zufallsexperiment mit genau zwei Ergebnissen* definiert, und es wird gesagt, dass eine Bernoulli-Kette aus mehreren *voneinander unabhängigen* Durchführungen eines Bernoulli-Experimentes besteht."

Der für die Bernoulli-Kette grundlegende Begriff der stochastischen Unabhängigkeit von Ereignissen tritt aber selbst für zwei Ereignisse in diesem Zusammenhang nicht mehr auf.

Die stiefmütterliche Behandlung der stochastischen Unabhängigkeit zeigt sich auch in der auf Seite 145 in Eichler und Vogel (2011) zu lesenden Definition:

> „Die n-malige *paarweise stochastische Wiederholung* eines Bernoulli-Experiments heißt Bernoulli-Kette der Länge n."

Abgesehen davon, dass stochastische Unabhängigkeit nur für Ereignisse als Teilmengen eines Grundraumes (Ergebnismenge) definiert ist, kommt man leider bei der Erzeugungsweise der Binomialverteilung nicht umhin, etwas zur Unabhängigkeit von n Ereignissen zu sagen.

Definition 2: n Ereignisse heißen (stochastisch) unabhängig, wenn für *jede Auswahl* von *mindestens zwei* dieser Ereignisse die Wahrscheinlichkeit des Schnittes der ausgewählten Ereignisse gleich dem Produkt der einzelnen Wahrscheinlichkeiten ist. Gilt diese „Schnitt-Produkt-Beziehung" nur für jeweils zwei der n Ereignisse, so liegt paarweise stochastische Unabhängigkeit vor.

Wohingegen in früheren Schulbüchern auch etwas zur Unabhängigkeit von mehr als zwei Ereignisse zu lesen war (siehe z. B. Felmy et al. (1986, S. 66), oder Barth und Haller (1985, S. 156)), findet sich heute nur noch die Unabhängigkeit von zwei Ereignissen (siehe Biehler et al. (2012, S. 75), oder Baum et al. (2007, S. 126)).

Dass aus paarweiser stochastischer Unabhängigkeit nicht die Unabhängigkeit von mehr als zwei Ereignissen folgt, zeigt das folgende instruktive

Beispiel 3: Wir betrachten die Ergebnismenge

$$\Omega := \{(0,0,0),(0,1,1),(1,0,1),(1,1,0)\}$$

und legen eine Gleichverteilung auf Ω zugrunde. Man beachte, dass die beiden ersten Komponenten der vier Tripel alle Ergebnis-Paare (a_1,a_2) mit $a_1,a_2 \in \{0,1\}$ durchlaufen und sich die jeweils letzte Komponente des Tripels deterministisch durch Addition modulo 2 von a_1 und a_2 ergibt. Definieren wir die Ereignisse

$$A_j := \{(a_1,a_2,a_3) \in \Omega : a_j = 1\}, \quad j = 1,2,3,$$

so gelten $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = 1/2$, da jedes dieser Ereignisse durch zwei Tripel gebildet wird. Weiter hat jeder der drei Schnitte von jeweils zweien dieser Ereignisse die Wahrscheinlichkeit $1/4$, denn er besteht immer aus genau einem Tripel. Somit sind A_1, A_2 und A_3 paarweise stochastisch unabhängig. Wegen $A_1 \cap A_2 \cap A_3 = \emptyset$ liegt jedoch keine Unabhängigkeit vor, denn dann müsste ja $\mathbb{P}(A_1 \cap A_2 \cap A_3) = 1/8$ gelten. $\square$

Steht eine 1 bzw. 0 für Treffer bzw. Niete, so lässt sich eine Bernoulli-Kette der Länge n mit Trefferwahrscheinlichkeit p durch den Grundraum

$$\Omega = \{(a_1, \ldots, a_n) : a_j \in \{0,1\} \text{ für } j = 1, \ldots, n\}$$

beschreiben. Weiter setzt man für $(a_1, \ldots, a_n) \in \Omega$

$$\mathbb{P}(\{(a_1, \ldots, a_n)\}) := p^{a_1+\ldots+a_n} \cdot (1-p)^{n-(a_1+\ldots+a_n)}.$$

Die Wahrscheinlichkeit eines n-Tupels $\omega = (a_1, \ldots, a_n)$ hängt also nur von der Anzahl $a_1 + \ldots + a_n$ seiner Einsen ab. Dieser Sachverhalt erschließt sich Schülerinnen und Schülern leicht durch die üblicherweise bis zur dritten Stufe gezeichneten Baumdiagramme, siehe z. B. Seite 130 in Freudigmann et al. (2016). Im obigen n-Tupel $(a_1, \ldots, a_n)$ steht a_j für einen Treffer (1) bzw. eine Niete (0) im j-ten Versuch, und das Ereignis

$$A_j := \{(a_1, \ldots, a_n) \in \Omega : a_j = 1\} \qquad (12)$$

beschreibt im obigen Grundraum Ω das aus 2^{n-1} Ergebnissen bestehende Ereignis „Treffer im j-ten Versuch“. Man beachte, dass durch die Verwendung der Symbole 1 und 0 für Treffer bzw. Niete die Anzahl der Treffer als Indikatorsumme der Form (1) geschrieben werden kann.

Wir werden jetzt zeigen, dass Schülerinnen und Schüler genügend Einsicht gewinnen können, dass im obigen Wahrscheinlichkeitsraum als Modell für eine Bernoulli-Kette der Länge n die durch (12) definierten Ereignisse $A_1, \ldots, A_n$ stochastisch unabhängig sind und die gleiche Wahrscheinlichkeit p besitzen.

Satz 2: In obigem Wahrscheinlichkeitsraum sind die Ereignisse $A_1, \ldots, A_n$ stochastisch unabhängig, und es gilt $\mathbb{P}(A_1) = \ldots = \mathbb{P}(A_n) = p$.

BEWEIS: Wir betrachten zunächst den Spezialfall $n = 4$, anhand dessen aber auch die allgemeine Argumentation klar wird. In diesem Fall ist der Grundraum durch

$$\Omega = \left\{ \begin{array}{llll} (0,0,0,0), & (0,0,0,1), & (0,0,1,0), & (0,0,1,1), \\ (0,1,0,0), & (0,1,0,1), & (0,1,1,0), & (0,1,1,1), \\ (1,0,0,0), & (1,0,0,1), & (1,0,1,0), & (1,0,1,1), \\ (1,1,0,0), & (1,1,0,1), & (1,1,1,0), & (1,1,1,1) \end{array} \right\}$$

gegeben. Wir setzen $q := 1 - p$ und betrachten zunächst das Ereignis A_1, das sich aus den in den beiden letzten Zeilen stehenden acht 4-Tupeln zusammensetzt. Die Wahrscheinlichkeit für jedes dieser Tupel enthält – weil an der ersten Stelle eine Eins steht – den Faktor p. Klammern wir diesen aus, so müssen wir die Wahrscheinlichkeiten für alle acht möglichen Tripel aus Einsen und Nullen addieren. Nach Zusammenfassen folgt

$$\begin{aligned} \mathbb{P}(A_1) &= p \cdot \left(q^3 + 3pq^2 + 3p^2q + p^3\right) \\ &= p \cdot (p+q)^3 = p. \end{aligned}$$

Das Argument kann aber nach dem Ausklammern von p noch einfacher sein: Die Summe der Wahrscheinlichkeiten der acht Tripel ist als Summe aller Ergebnisse einer Bernoulli-Kette der Länge drei gleich Eins! Hieran erkennt man schon anhand eines Symmetrieargumentes, dass auch gilt:

$$\mathbb{P}(A_2) = \mathbb{P}(A_3) = \mathbb{P}(A_4) = p$$

Das Ereignis $A_1 \cap A_2$ setzt sich aus den in der unteren Zeile stehenden vier Quadrupeln zusammen. Die Wahrscheinlichkeit für jedes dieser Quadrupel enthält wegen der beiden jeweils zu Beginn stehenden Einsen den „ausklammerbaren“ Faktor p^2, und man erhält

$$\begin{aligned} \mathbb{P}(A_1 \cap A_2) &= p^2 \cdot \left(q^2 + 2pq + p^2\right) = p^2 \cdot (p+q)^2 \\ &= p^2 = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2). \end{aligned}$$

Die Ereignisse A_1 und A_2 sind also stochastisch unabhängig. Wiederum aus Symmetriegründen (bei den vier das Ereignis $A_i \cap A_j$ für $i \neq j$ bildenden Quadrupeln stehen die ausklammerbaren Einsen an den Stellen i und j, und nach Ausklammern summieren sich die Wahrscheinlichkeiten einer Bernoulli-Kette der Länge 2 zu Eins) folgt, dass allgemein

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j), \quad 1 \leq i < j \leq n,$$

gilt, also paarweise stochastische Unabhängigkeit vorliegt. Wenn wir jetzt das aus nur zwei 4-Tupeln bestehende Ereignis $A_1 \cap A_2 \cap A_3$ betrachten, folgt

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = p^3 \cdot (p+q) = \prod_{j=1}^{3} \mathbb{P}(A_j),$$

und wiederum aus Symmetriegründen ist die Wahrscheinlichkeit jedes Schnittes von drei der vier Ereignisse gleich dem Produkt der einzelnen Wahrscheinlichkeiten. Zu guter Letzt gilt

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = p^4 = \prod_{j=1}^{4} \mathbb{P}(A_j),$$

was insbesondere zeigt, dass $A_1, \ldots, A_4$ stochastisch unabhängig sind und die gleiche Wahrscheinlichkeit p besitzen.

Die allgemeine Argumentation für eine aus allen 2^n n-Tupeln aus Einsen und Nullen einer Bernoulli-

Kette der Länge n erfordert keinen Mehraufwand: Wählt man für ein k mit $1 \le k \le n-1$ die Stellen $i_1,\ldots,i_k$ mit $1 \le i_1 < \ldots < i_k \le n$ im Tupel aus, so kann bei der Berechnung von $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})$ (für $k = 1$ entsteht kein Schnitt) wegen der jeweils an den Stellen $i_1,\ldots,i_k$ stehenden Einsen der Faktor p^k ausgeklammert werden. Innerhalb der Klammer steht die Summe über die Wahrscheinlichkeiten aller $n-k$-Tupel der Ergebnisse einer Bernoulli-Kette der Länge $n-k$, und diese ist gleich 1. Also ergibt sich

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = p^k = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).$$

Da auch $\mathbb{P}(A_1 \cap \ldots \cap A_n) = p^n$ gilt, sind $A_1,\ldots,A_n$ stochastisch unabhängig. □

7 Die Binomialverteilung – Schlüssel*kon*(*re*?)zept?

In Freudigmann et al. (2016) wird die Binomialverteilung als „Schlüsselkonzept“ hervorgehoben. Auf 30 Seiten findet man dann aber doch im Wesentlichen nur Schlüssel*rezepte*. Die Begriffsbildung *stochastische Unabhängigkeit* tritt nicht auf, und selbst wiederholt geschossene Elfmeter oder Basketballfreiwürfe werden kommentarlos als Bernoulli-Kette deklariert; die Anzahlen der jeweils erzielten Treffer sind selbstverständlich binomialverteilt. Für eine Zufallsvariable X mit der Binomialverteilung $\mathrm{Bin}(n,p)$ finden Schülerinnen und Schüler die zentrale Formel

$$\mathbb{P}(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}, \qquad (13)$$

$k = 0,1,\ldots,n$, von Bernoulli, und es folgen Übungsaufgaben, in denen aus der Gleichung

$$\mathbb{P}(X = 4) = \binom{6}{4} \cdot \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^2$$

auf n, k und p geschlossen werden soll. Auf Seite 140 erfährt man, dass der Erwartungswert von X gleich $n \cdot p$ ist, wobei der Nachweis dieser Beziehung im Fall $n = 3$ erfolgt. Auf der nächsten Seite finden sich Übungsaufgaben der Gestalt, dass man in den Fällen a) $n = 10, p = 0{,}2$, b) $n = 20, p = 0{,}7$, c) $n = 50, p = 2/3$ und d) $n = 90, p = 1/6$ den Erwartungswert ausrechnen soll. Hier ist kein Denken mehr gefragt, sondern simple „Formelbefolgung“.

Auch die Denkarbeit, denjenigen Wert von k zu bestimmen, für den die Wahrscheinlichkeit $\mathbb{P}(X = k)$ maximal wird, wird Schülerinnen und Schülern abgenommen. Auf Seite 140 steht hierzu die Bemerkung: „Falls $n \cdot p$ ganzzahlig ist, hat diese Trefferanzahl die höchste Wahrscheinlichkeit und das entsprechende Rechteck im Histogramm ist das höchste. Falls $n \cdot p$ nicht ganzzahlig ist, hat eine der beiden benachbarten Trefferanzahlen die maximale Wahrscheinlichkeit“. Auf der nächsten Seite gibt es dann Aufgaben, in denen dieses Rezept angewendet werden soll.

Es ist erhellend, wie früher (siehe z. B. Griesel et al. (2003, S. 118)) den Quotienten

$$\begin{aligned}\frac{\mathbb{P}(X = k)}{\mathbb{P}(X = k-1)} &= \frac{\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}}{\binom{n}{k-1} \cdot p^{k-1} \cdot (1-p)^{n-(k-1)}} \\ &= \frac{p}{1-p} \cdot \frac{n-k+1}{k}\end{aligned}$$

$(k = 1,2,\ldots,n)$ zu bilden und diesen mit 1 zu vergleichen: Es gilt

$$\frac{\mathbb{P}(X = k)}{\mathbb{P}(X = k-1)} \left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} 1 \iff k \left\{\begin{matrix} < \\ = \\ > \end{matrix}\right\} np + p.$$

Hieraus folgt durch Unterscheidung der Fälle $n \cdot p \in \mathbb{N}$ und $n \cdot p \notin \mathbb{N}$ die Behauptung.

Studierende des gymnasialen Lehramts Mathematik sollten im Zusammenhang mit der Binomialverteilung unbedingt folgendes Resultat verinnerlicht haben:

Satz 3: Sind $A_1,\ldots,A_n$ unabhängige Ereignisse mit gleicher Wahrscheinlichkeit p, so besitzt die Anzahl $X := \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \ldots + \mathbf{1}_{A_n}$ der eintretenden Ereignisse die Binomialverteilung $\mathrm{Bin}(n,p)$.

BEWEIS: Der Beweis verwendet, dass die Unabhängigkeit von Ereignissen erhalten bleibt, wenn man zu komplementären Ereignissen übergeht. Das Ereignis $\{X = k\}$ tritt genau dann ein, wenn es *irgendeine* Teilmenge T der Menge $\{1,2,\ldots,n\}$ mit genau k Elementen gibt, sodass alle Ereignisse A_i mit $i \in T$ eintreten und alle Ereignisse A_j mit $j \notin T$ *nicht* eintreten. Die Wahrscheinlichkeit, dass alle A_i mit $i \in T$ eintreten und die A_j mit $j \notin T$ nicht eintreten, ist wegen der stochastischen Unabhängigkeit gleich

$$\prod_{i \in T} \mathbb{P}(A_i) \cdot \prod_{j \notin T} \mathbb{P}(\overline{A}_j).$$

Da T k Elemente besitzt und $\mathbb{P}(A_i) = p$ sowie $\mathbb{P}(\overline{A}_j) = 1-p$ gelten, ist obiges Produkt gleich $p^k \cdot (1-p)^{n-k}$. Bedenkt man noch, dass es $\binom{n}{k}$ k-elementige Teilmengen der Menge $\{1,\ldots,n\}$ gibt, so folgt (13). □

Anhand dieses Resultates wird klar, dass man gar nicht ein und dasselbe Bernoulli-Experiment in unabhängiger Folge wiederholen muss, damit die Anzahl der Treffer binomialverteilt ist.

So könnte man etwa im ersten Experiment eine echte Münze werfen und „Zahl“ als Treffer deklarieren, in einem zweiten Versuch einen echten Würfel werfen, wobei ein Treffer das Auftreten von 2, 4, oder 6 darstellt, und in einem dritten Experiment ein Glücksrad mit zwei gleich großen Sektoren drehen, von denen einer mit „Treffer“ beschriftet ist.

Es kommt also nur auf stochastische Unabhängigkeit und gleiche Trefferwahrscheinlichkeit an!

Beispiel 3 zeigt zudem eindrucksvoll, dass paarweise unabhängige Ereignisse mit gleicher Wahrscheinlichkeit nicht ausreichen, damit die Anzahl der eintretenden Ereignisse binomialverteilt ist. So kann in jenem Beispiel die Indikatorsumme $X = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \mathbf{1}_{A_3}$ nicht den Wert 3 annehmen.

8 Standardabweichung – mehr als nur eine Formel?

In Freudigmann et al. (2016) wird auf Seite 147 die Standardabweichung einer Verteilung wie folgt eingeführt:

> „Für eine beliebige Wahrscheinlichkeitsverteilung ist festgelegt:“
>
> $$\sigma = \sqrt{\begin{aligned}&(x_1-\mu)^2\mathbb{P}(X=x_1)+\ldots\\ &\quad\ldots+(x_n-\mu)^2\mathbb{P}(X=x_n).\end{aligned}}$$

Direkt danach erfährt man (ohne jeden Beweis), dass es bei der Binomialverteilung wie beim Erwartungswert auch eine einfache Formel für die Standardabweichung gebe, nämlich $\sigma = \sqrt{n \cdot p \cdot (1-p)}$. Ohne jegliches weiteres Hintergrundwissen über den zentralen Grenzwertsatz von de Moivre–Laplace findet man noch die folgende mathematisch falsche Aussage:

> „Man kann zeigen: Wenn bei einer Binomialverteilung die Parameter n und p genügend groß sind, beträgt die Wahrscheinlichkeit, dass die Trefferanzahl um höchstens σ vom Erwartungswert abweicht, ca. 68% “.

Der Fehler liegt hier in der Formulierung „und p genügend groß“. Bei *festem* p mit $0 < p < 1$ konvergiert eine Folge standardisierter Binomialverteilungen bei Grenzübergang $n \to \infty$ gegen die Standardnormalverteilung. Dabei ist die Approximation bei gegebenem n prinzipiell umso schlechter, je näher p bei Null oder bei Eins liegt; in Büchern zur Angewandten Statistik (siehe z. B. Stahel (1999, S. 154)) wird üblicherweise die Faustregel formuliert, dass die Normalverteilungsapproximation im Fall $n \cdot p \cdot (1-p) \geq 9$ akzeptabel sei.

In Freudigmann et al. (2016) wird dann ausschließlich für konkrete Fälle von n und p die Standardabweichung ausgerechnet und das obige „68%-Rezept“ befolgt.

Die erste Frage, die sich stellt, ist natürlich, was Schülerinnen und Schüler hiermit wirklich begreifen sollen? Kann man nicht auch die Standardabweichung weglassen, wenn es schon keine allgemeinen Einsichten über die Tschebyschev-Ungleichung und den Zentralen Grenzwertsatz von de Moivre–Laplace mehr gibt?

Im Folgenden möchte ich zwei Wege aufzeigen, wie man von Darstellung (5) und Satz 1 ausgehend die Standardabweichung einer binomialverteilten Zufallsvariablen herleiten kann, sodass zumindest Schülerinnen und Schüler eines zukünftigen Leistungskurses Einsichten vermittelt und Begründungen geliefert werden.

Betrachtet man analog zu (4) den Ausdruck

$$(X(\omega_1)-\mu)^2 \cdot \frac{h_1}{n} + \ldots + (X(\omega_k)-\mu)^2 \cdot \frac{h_k}{n},$$

also den *durchschnittlichen Wert der quadratischen Abweichung von X um den Erwartungswert $\mu = \mathbb{E}(X)$* pro Vorgang, so ist die nachfolgende Definition aufgrund der Stabilisierung der relativen Häufigkeiten h_j/n gegen $\mathbb{P}(\{\omega_j\})$ motiviert.

Definition 3: Ist X eine Zufallsvariable mit Erwartungswert $\mu = \mathbb{E}(X)$, so heißt die Zahl

$$\mathbb{V}(X) := \mathbb{E}\left[(X-\mu)^2\right]$$

Varianz von X. Die (nichtnegative) Wurzel aus $\mathbb{V}(X)$ heißt *Standardabweichung* von X.

Als Erwartungswert der Zufallsvariablen $(X-\mu)^2$ gilt wegen $(X-\mu)^2(\omega) = (X(\omega)-\mu)^2$, $\omega \in \Omega$, nach (5) definitionsgemäß

$$\mathbb{V}(X) = \sum_{j=1}^{k} (X(\omega_j)-\mu)^2 \cdot \mathbb{P}(\{\omega_j\}).$$

Nimmt X die Werte $x_1, \ldots, x_s$ an, so folgt hieraus durch Zusammenfassen nach gleichen Werten von $X(\omega_j)$ die Gleichung

$$\mathbb{V}(X) = \sum_{\ell=1}^{s} (x_\ell-\mu)^2 \cdot \mathbb{P}(X = x_\ell)$$

und nach Ziehen der Wurzel die zu Beginn dieses Abschnitts gegebene Form für σ, wobei nur n durch s zu ersetzen ist.

Da nach Definition 3 die Varianz ein *Erwartungswert* ist (von der quadratischen Abweichung der Zufallsvariablen X um ihren Erwartungswert), und da die Erwartungswertbildung nach Satz 1 gewissen Rechenregeln genügt, erhält man folgende alternative Darstellungen der Varianz:

Satz 4: Mit $\mu := \mathbb{E}(X)$ gilt

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}(X^2) - \mu^2 \\ &= \mathbb{E}\big[X \cdot (X-1)\big] + \mu - \mu^2.\end{aligned}$$

BEWEIS: Aus Satz 1 folgt

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}\big(X^2 - 2\mu \cdot X + \mu^2\big) \\ &= \mathbb{E}(X^2) - 2\mu \cdot \mathbb{E}(X) + \mu^2 \\ &= \mathbb{E}(X^2) - \mu^2 \\ &= \mathbb{E}\big[X \cdot (X-1) + X\big] - \mu^2 \\ &= \mathbb{E}\big[X \cdot (X-1)\big] + \mu - \mu^2,\end{aligned}$$

was zu zeigen war. □

Mit der binomischen Formel ergibt sich jetzt recht leicht die Varianz (und damit auch die Standardabweichung) der Binomialverteilung.

Satz 5: Eine $\mathrm{Bin}(n,p)$-verteilte Zufallsvariable X besitzt die Varianz $n \cdot p \cdot (1-p)$.

BEWEIS: Es gilt

$$\mathbb{E}\big[X \cdot (X-1)\big] = \sum_{k=0}^{n} k(k-1) \cdot \binom{n}{k} p^k (1-p)^{n-k}.$$

Da die Summation ab $k = 2$ erfolgen kann und im Fall $k \geq 2$

$$k(k-1) \cdot \binom{n}{k} = n(n-1) \cdot \binom{n-2}{k-2}$$

gilt, so folgt nach Abspaltung von p^2 mit der Abkürzung $a = \mathbb{E}[X(X-1)]$

$$a = n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} \cdot (1-p)^{n-2-(k-2))}.$$

Ersetzen wir hier $k-2$ durch j, so ergibt sich

$$a = n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j}.$$

Nach dem binomischen Lehrsatz ist die Summe gleich $(p+1-p)^{n-2} = 1$, und es folgt mit Satz 4

$$\begin{aligned}\mathbb{V}(X) &= n \cdot (n-1) \cdot p^2 + n \cdot p - (n \cdot p)^2 \\ &= n \cdot p \cdot (1-p).\end{aligned}$$

Eine andere Möglichkeit, die Varianz einer binomialverteilten Zufallsvariablen zu bestimmen, liefert folgendes allgemeine Resultat über die Varianz einer Indikatorsumme.

Satz 6: Ist $X = \mathbf{1}_{A_1} + \ldots + \mathbf{1}_{A_n}$ eine Indikatorsumme, so gilt

$$\begin{aligned}\mathbb{V}(X) &= \sum_{i=1}^{n} \mathbb{P}(A_i) \cdot (1 - \mathbb{P}(A_i)) \\ &\qquad + \sum_{i \neq j} \big(\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)\big).\end{aligned}$$

BEWEIS: Wir verwenden (6) sowie die Darstellung $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. Es gilt

$$\begin{aligned}X^2 &= \left(\sum_{i=1}^{n} \mathbf{1}_{A_i}\right) \cdot \left(\sum_{j=1}^{n} \mathbf{1}_{A_j}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{1}_{A_i} \cdot \mathbf{1}_{A_j} = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{1}_{A_i \cap A_j}\end{aligned}$$

und somit nach Satz 1 a) und c)

$$\mathbb{E}(X^2) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{P}(A_i \cap A_j).$$

Wegen

$$(\mathbb{E}X)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$$

folgt durch Differenzbildung und Aufspaltung der Doppelsumme nach den Summanden mit $i = j$ und $i \neq j$ die Behauptung. □

Da im Fall einer binomialverteilten Zufallsvariablen die Ereignisse $A_1, \ldots, A_n$ stochastisch unabhängig sind, verschwindet jeder der Summanden in der in Satz 6 auftretenden Doppelsumme. Da zudem $A_1, \ldots, A_n$ die gleiche Wahrscheinlichkeit p besitzen, liefert Satz 6 unmittelbar das Resultat $\mathbb{V}(X) = n \cdot p \cdot (1-p)$. Satz 6 zeigt auch, dass die Varianz einer Indikatorsumme X nur von den Wahrscheinlichkeiten der beteiligten Ereignisse und der Wahrscheinlichkeit der Schnitte von je zwei Ereignissen abhängt. Man benötigt also nicht die Verteilung von X, um die Varianz von X zu erhalten!

Beispiel 4: (Fortsetzung von Beispiel 1)
Wir betrachten die in (8) definierten Ereignisse A_1, A_2, A_3, A_4 und behaupten, dass diese stochastisch unabhängig sind. Damit wäre nach den Ergebnissen von Beispiel 1 und Satz 6 die Varianz der Anzahl X der Rekorde durch

$$\mathbb{V}(X) = \sum_{j=1}^{4} \frac{1}{j}\left(1 - \frac{1}{j}\right) = \frac{95}{144}$$

gegeben. Schülerinnen und Schüler würden hier natürlich durch Abzählen der jeweils günstigen Fälle vorgehen, wobei schnell klar wird, dass das Ereignis A_1 wegen $A_1 = \Omega$ „außen vor bleiben kann“, denn sein Schnitt mit anderen Ereignissen hat keinen Einfluss, und seine Wahrscheinlichkeit ist gleich 1. Das Ereignis $A_2 \cap A_3$ besteht aus den vier Permutationen 1234, 1243, 1342 und 2341. Somit gilt $\mathbb{P}(A_1 \cap A_2) = 1/6 = 1/2 \cdot 1/3 = \mathbb{P}(A_2) \cdot \mathbb{P}(A_3)$ usw.

Hier sollte die Lehrkraft aber auch Einsicht vermitteln, *warum* ganz allgemein in einer rein zufälligen Permutation der Zahlen $1, 2, \ldots, n$ die Ereignisse A_i und A_j für $2 \leq i < j \leq n$ stochastisch unabhängig sind. Nun, das Eintreten von A_j bedeutet, dass von links gelesen die j-te Zahl die größte der ersten j Zahlen ist. Diese Information hat keinen Einfluss auf die Reihenfolge der ersten i Zahlen. Irgendeine davon ist die größte, und dass diese Zahl an der i-ten Stelle steht, ist aus Symmetriegründen gleich $1/i$. Die bedingte Wahrscheinlichkeit von A_i unter der Bedingung A_j ist also gleich $\mathbb{P}(A_i)$, und somit sind A_i und A_j stochastisch unabhängig. □

Auch die früher in Schulbüchern behandelte Tschebyschev-Ungleichung kann ohne größere Schwierigkeiten erhalten werden:

Satz 7: (Tschebyschev–Ungleichung)
Seien X eine Zufallsvariable mit Erwartungswert μ und $\varepsilon > 0$ eine beliebige Zahl. Dann gilt:

$$\mathbb{P}\left(|X - \mu| \geq \varepsilon\right) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

BEWEIS: Sei $A := \{\omega \in \Omega : |X(\omega) - \mu| \geq \varepsilon\}$. Für jedes $\omega \in \Omega$ gilt

$$\mathbf{1}_A(\omega) \ \leq \ \frac{(X(\omega) - \mu)^2}{\varepsilon^2},$$

da im Fall $\omega \in A$ die rechte Seite der obigen Ungleichung größer oder gleich Eins ist. Mit Satz 1 c), d) und b) folgt

$$\begin{aligned} \mathbb{P}(|X - \mu| \geq \varepsilon) \ &= \ \mathbb{P}(A) \ = \ \mathbb{E}\left(\mathbf{1}_A\right) \\ &\leq \ \mathbb{E}\left(\frac{1}{\varepsilon^2} \cdot (X - \mu)^2\right) \\ &= \ \frac{\mathbb{E}\left((X - \mu)^2\right)}{\varepsilon^2} = \frac{\mathbb{V}(X)}{\varepsilon^2}. \ \square \end{aligned}$$

9 Schlussbemerkungen

Der gymnasiale Stochastikunterricht ist momentan nicht dazu angetan, Schülerinnen und Schülern Einsichten in die Wirkungsweise des Zufalls zu vermitteln. Insbesondere im Zusammenhang mit Bernoulli-Ketten und der Binomialverteilung ist das, was vermittelt wird, zu einfachen Rezepten degeneriert. Am augenfälligsten wird diese „Rezepteritis“ anhand der Standardabweichung der Binomialverteilung und der völlig vom Himmel fallenden *68%-Regel*. Durch eine solche Unterrichtskultur wird die schon große Kluft zwischen der Mathematik im Gymnasium und der universitären Mathematik noch weiter vertieft.

In dieser Arbeit wurde eine sich an den Grundbegriffen Zufallsvariable, Erwartungswert, Binomialkoeffizient, binomische Formel, Bernoulli-Kette, Binomialverteilung sowie Varianz und Standardabweichung orientierende Bestandsaufnahme vorgenommen und gezeigt, dass man allein durch Einführung des Erwartungswertes auf der Ergebnismenge eines stochastischen Vorgangs über die Binomialverteilung hinausgehende Einsichten gewinnen kann.

Als Fachstochastiker plädiere ich vehement dafür, weniger Stochastik in der Schule zu behandeln, das Weniger dann aber unter der Prämisse des Begreifens und Verstehens.

Danksagung: Ich danke Daniel Hug, Reimund Vehling sowie dem Schulgutachter für eine gründliche Durchsicht des Manuskriptes und hilfreiche Hinweise.

Literatur

Baum, M. et al. (2007): Lambacher Schweizer 5. Mathematik für Gymnasien Baden–Württemberg. Stuttgart: Klett.

Barth, F.; Haller, R. (1985): Stochastik Leistungskurs. München: Ehrenwirth.

Biehler, R. et al. (2012): Mathematik Neue Wege. Arbeitsbuch für Gymnasien, Stochastik. Braunschweig: *Schroedel*.

Bostelmann, M. et al. (2009): Mathematik Neue Wege 6. Arbeitsbuch für Gymnasien Baden-Württemberg. Braunschweig: *Schroedel*.

Dürr, R. et al. (2017): Lambacher Schweizer Mathematik für Gymnasien. Kursstufe. Stuttgart: Klett.

Eichler, A.; Vogel, M. (2011): Leitfaden Stochastik: Für Studierende und Ausübende des Lehramts. Wiesbaden: Vieweg+Teubner.

Felmy, H.-G.; Schmidt, A.; Stark, J. (1986): LS Mathematik. Stochastik Grundkurs. Stuttgart: Klett.

Freudigmann, H., et al. (2016): Lambacher Schweizer 10. Mathematik für Gymnasien Baden–Württemberg. Stuttgart: Klett.

Glaser, H. et al. (1990). Sigma: Grundkurs Stochastik. Stuttgart: Klett.

Griesel, H. et al. (2003): Elemente der Mathematik. Leistungskurs Stochastik. Hannover: Schroedel.

Henze, N. (2008): Rekorde. In: *Der Mathematikunterricht*, 54, S. 16–23.

Henze, N. (2017): Stochastik für Einsteiger. 11. Auflage: Heidelberg: Springer Spektrum.

Henze, N. (o.D.): Einführung in die Stochastik für Studierende des gymnasialen Lehramts Mathematik. Karlsruhe: Karlsruher Institut für Technologie. www.youtube.com/watch?v=CXSyMvzl8cY&list=PLtexvciJ0k3FWWdVjNlItYuXHhRn1Qpf (Zugriff: 26.7.2018)

Stahel, W. (1999): Statistische Datenanalyse. Braunschweig: Vieweg.

Anschrift des Verfassers

Prof. Dr. Norbert Henze
Institut für Stochastik
Karlsruher Institut für Technologie (KIT)
Englerstr. 2
76131 Karlsruhe
Henze@kit.edu